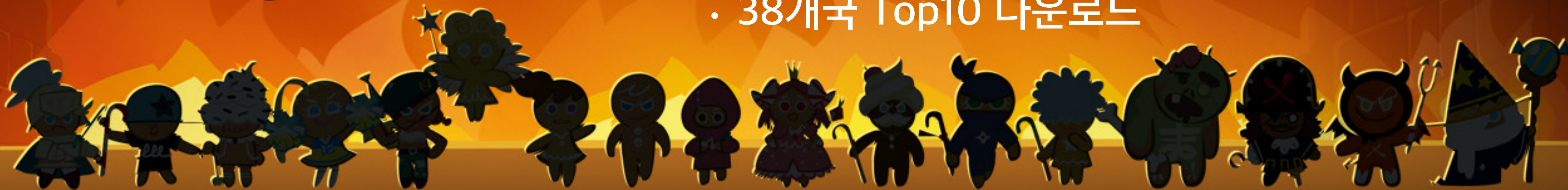
AWS를 활용하여 Daily Report 만들기

로그 수집부터 자동화된 분석까지

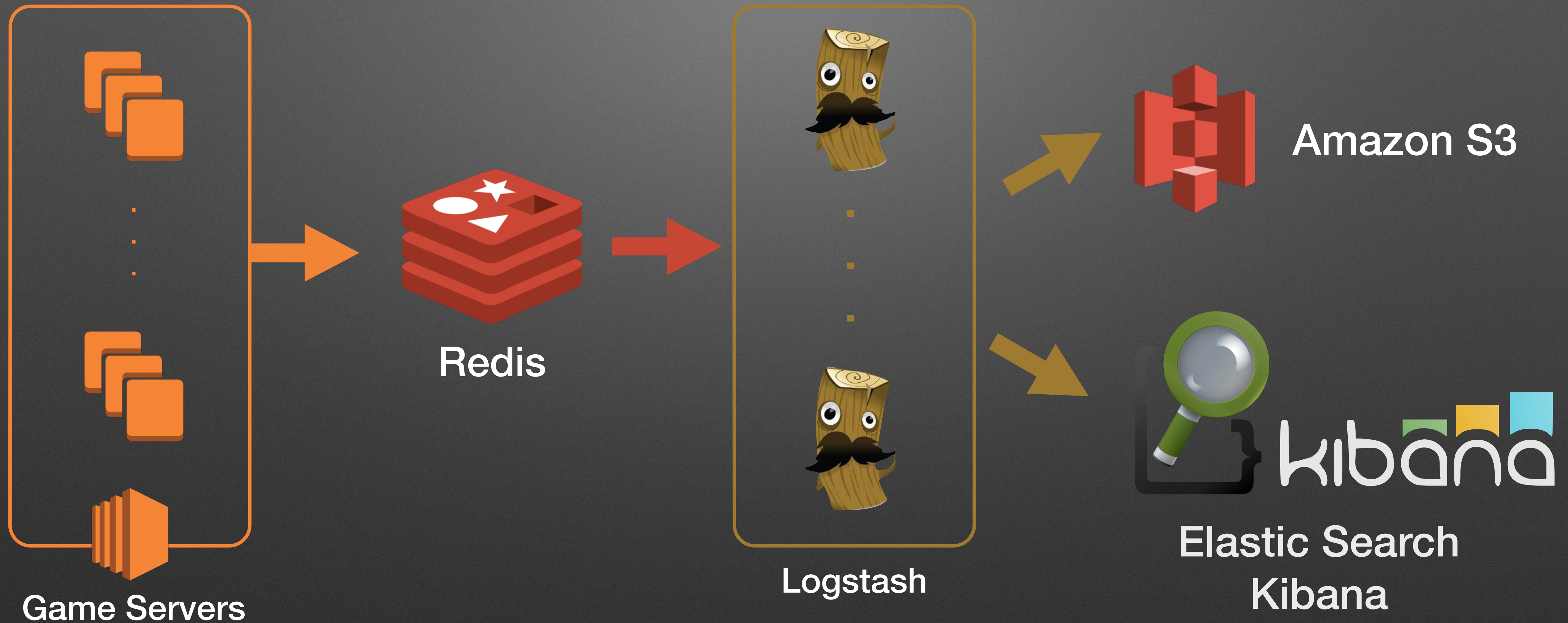
서버개발팀
정창제
changje@devsisters.com

쿠키런 소개

- 7,000만 누적 다운로드
- 최대 1,000만 DAU
- 한국, 일본, 태국 등 10여개국 다운로드 1위
- 38개국 Top10 다운로드

쿠키런의 하루 로그는 200~400GB
1년이면 100TB

차곡차곡 잘 쌓아야 한다

0. 로그 수집

Game Servers
Redis
Logstash
Amazon S3
kibana
Elastic Search
Kibana

쿠키런의 하루 로그는 200~400GB
1년이면 100TB

차곡차곡 잘 쌓아야 한다

차곡차곡 쌓여가는 로그 하지만 분석할 방법이 없다..?

쿠키런의 하루 로그는 200~400GB
1년이면 100TB

차곡차곡 잘 쌓아야 한다

차곡차곡 쌓여가는 로그 하지만 분석할 방법이 없다..?

최적의 분석시스템을 만들자!

BREADCRUMB

1. 쉽고 빠르게 로그 분석하기

1. 로그분석시스템 요구조건

- 하루 전 로그와 1년 전 로그를 동일하게!

→ 날짜별로 로그파일을 분리

1. 로그분석시스템 요구조건

- 하루 전 로그와 1년 전 로그를 동일하게!

  → 날짜별로 로그파일을 분리

- 특정 활동 및 특정 유저의 로그를 빠르게!

  → Action과 MemberID로 Index

1. 로그분석시스템 요구조건

- 하루 전 로그와 1년 전 로그를 동일하게!

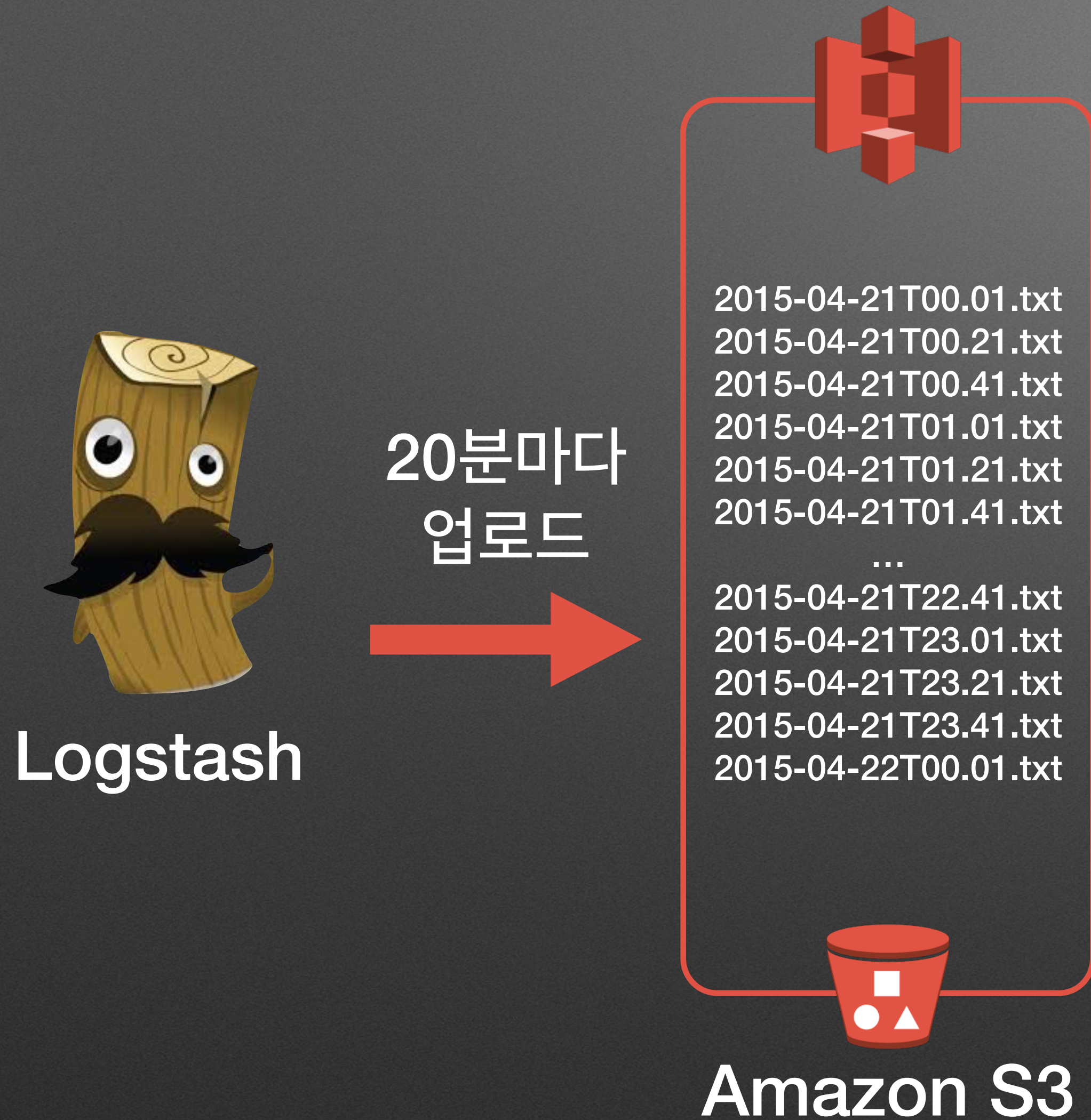
  → 날짜별로 로그파일을 분리

- 특정 활동 및 특정 유저의 로그를 빠르게!

  → Action과 MemberID로 Index

- 로컬 개발환경에서도 빠르게!

  → 유저 샘플링을 지원하자!

1. 로그 전처리

Logstash

20분마다
업로드

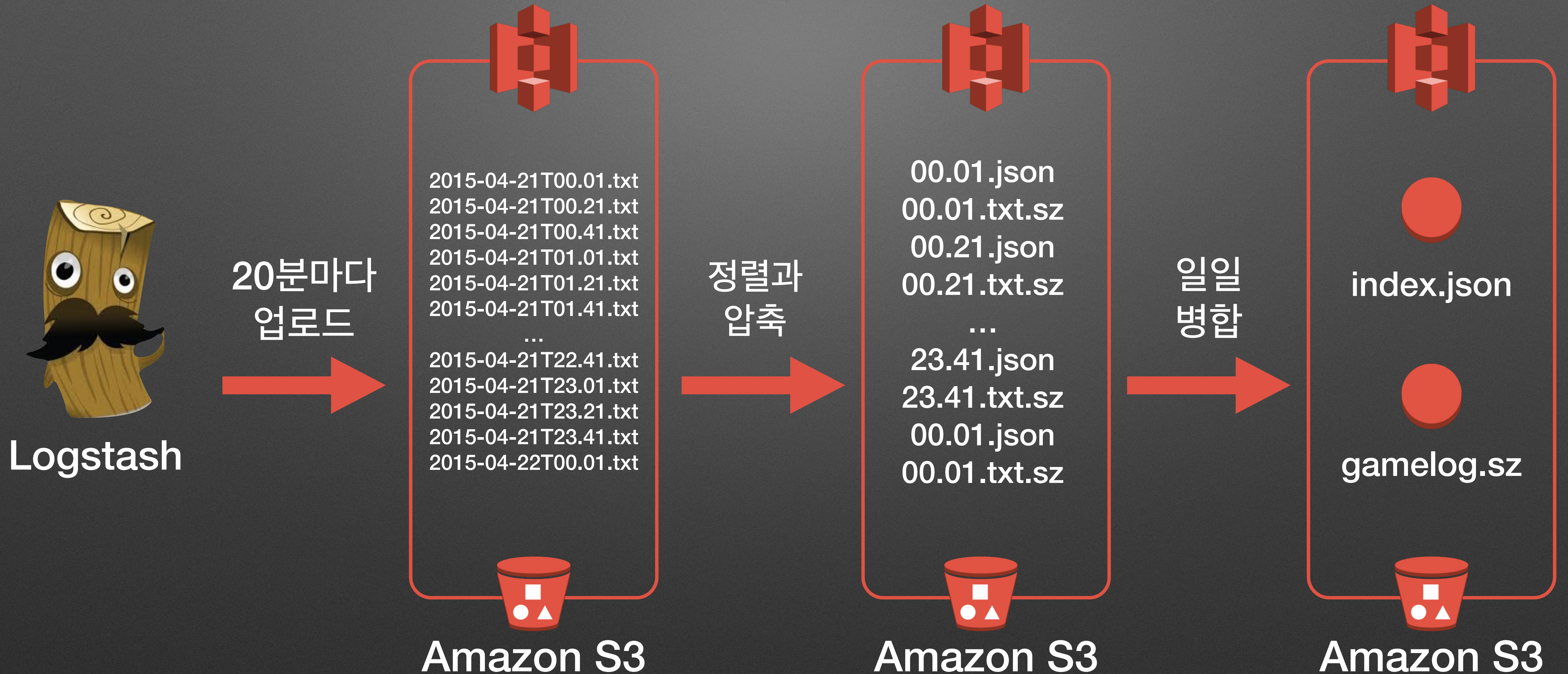
2015-04-21T00.01.txt
2015-04-21T00.21.txt
2015-04-21T00.41.txt
2015-04-21T01.01.txt
2015-04-21T01.21.txt
2015-04-21T01.41.txt
...
2015-04-21T22.41.txt
2015-04-21T23.01.txt
2015-04-21T23.21.txt
2015-04-21T23.41.txt
2015-04-22T00.01.txt

Amazon S3

1. 로그 전처리

Logstash
20분마다
업로드
2015-04-21T00.01.txt
2015-04-21T00.21.txt
2015-04-21T00.41.txt
2015-04-21T01.01.txt
2015-04-21T01.21.txt
2015-04-21T01.41.txt
...
2015-04-21T22.41.txt
2015-04-21T23.01.txt
2015-04-21T23.21.txt
2015-04-21T23.41.txt
2015-04-22T00.01.txt
Amazon S3
정렬과
압축
00.01.json
00.01.txt.sz
00.21.json
00.21.txt.sz
...
23.41.json
23.41.txt.sz
00.01.json
00.01.txt.sz
Amazon S3

1. 로그 전처리

Logstash
20분마다
업로드
2015-04-21T00.01.txt
2015-04-21T00.21.txt
2015-04-21T00.41.txt
2015-04-21T01.01.txt
2015-04-21T01.21.txt
2015-04-21T01.41.txt
...
2015-04-21T22.41.txt
2015-04-21T23.01.txt
2015-04-21T23.21.txt
2015-04-21T23.41.txt
2015-04-22T00.01.txt
Amazon S3
정렬과
압축
00.01.json
00.01.txt.sz
00.21.json
00.21.txt.sz
...
23.41.json
23.41.txt.sz
00.01.json
00.01.txt.sz
Amazon S3
일일
병합
index.json
gamelog.sz
Amazon S3

1. 로그 전처리

gamelog.sz

- 날짜로 분리
- Action, MemberID로 정렬
  - 특정 활동 및 특정 유저의 로그를 빠르게 찾을 수 있도록
  - 로컬 개발에서 샘플링하여 사용할 수 있도록 MemberID를 뒤집은 값을 키로 사용
- Block 단위 Snappy 압축
  - Networt Latency ↓
  - Snappy : CPU 사용량 ↓
  - Splittable (Snappy Framing Format)

1. 로그 전처리

index.json

- gamelog를 조회하기 위한 Metadata

```
…},
{
  "key_first": “<Action> <Reversed MemberID> <Time> <Hash>”,
  "key_last": “<Action> <Reversed MemberID> <Time> <Hash>”,
  "offset": 0,
  "length": <Length>
}, …
```

1. 로그 전처리

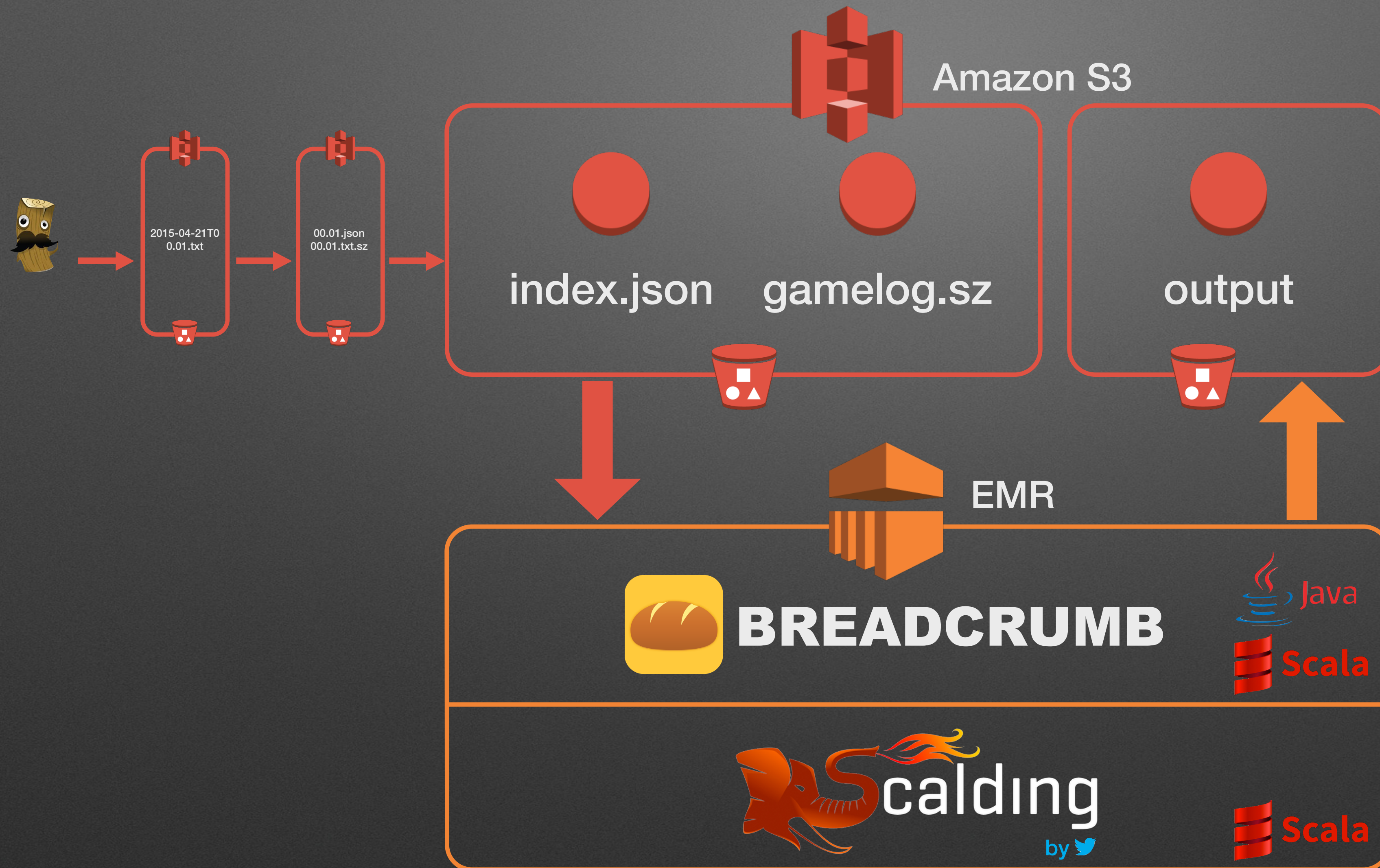
gamelog.sz

member/connect	10000
member/connect	240000
member/connect	250000
member/disconnect	3000
member/disconnect	13000
game/startGame	00012

index.json

[{
  "key_first": “member/connect 00001”,
  "key_last": “member/connect 000024”,
  "offset": 0,
  “length": 31243
}, {
  "key_first": “member/connect 000025”,
  "key_last": “member/disconnect 0003”,
  "offset": 31244,
  "length": 42347
}, {
  "key_first": “member/disconnect 00031”,
  "key_last": “game/startGame 00012”,
  "offset": 73591,
  "length": 49323
}]

1. 로그 분석

https://github.com/twitter/scalding

1. 예제 - ActiveUser

```
class ActiveUserCount(args : Args) extends Job(args) {
  CookierunLog("member/connect", args("sampling"),
    args("timeStart"), args("timeEnd"),
    Tuple1("memberSeq"))
  .groupBy(‘memberSeq){ _.take(1) }
  .groupAll{ _.size }
  .write(Csv(args("output")))
}
```

1/1000 샘플링시 1주일에 대해 로컬에서 15초

https://github.com/twitter/scalding

로그도 쌓이고...

분석 시스템도 있는데...

요청이 있을때만 비정기적으로 추출하는 통계...

로그도 쌓이고...

분석 시스템도 있는데...

요청이 있을때만 비정기적으로 추출하는 통계...

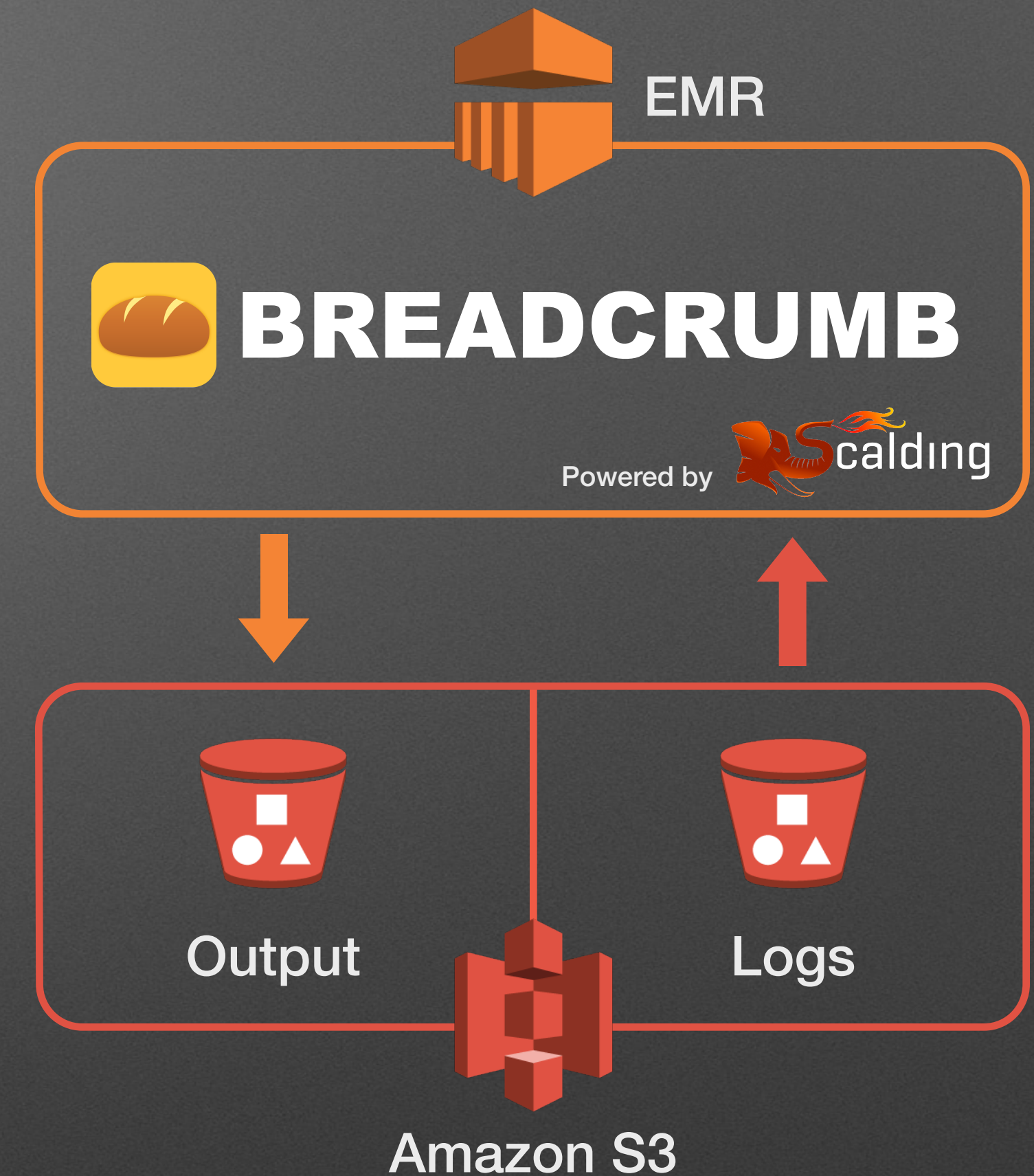
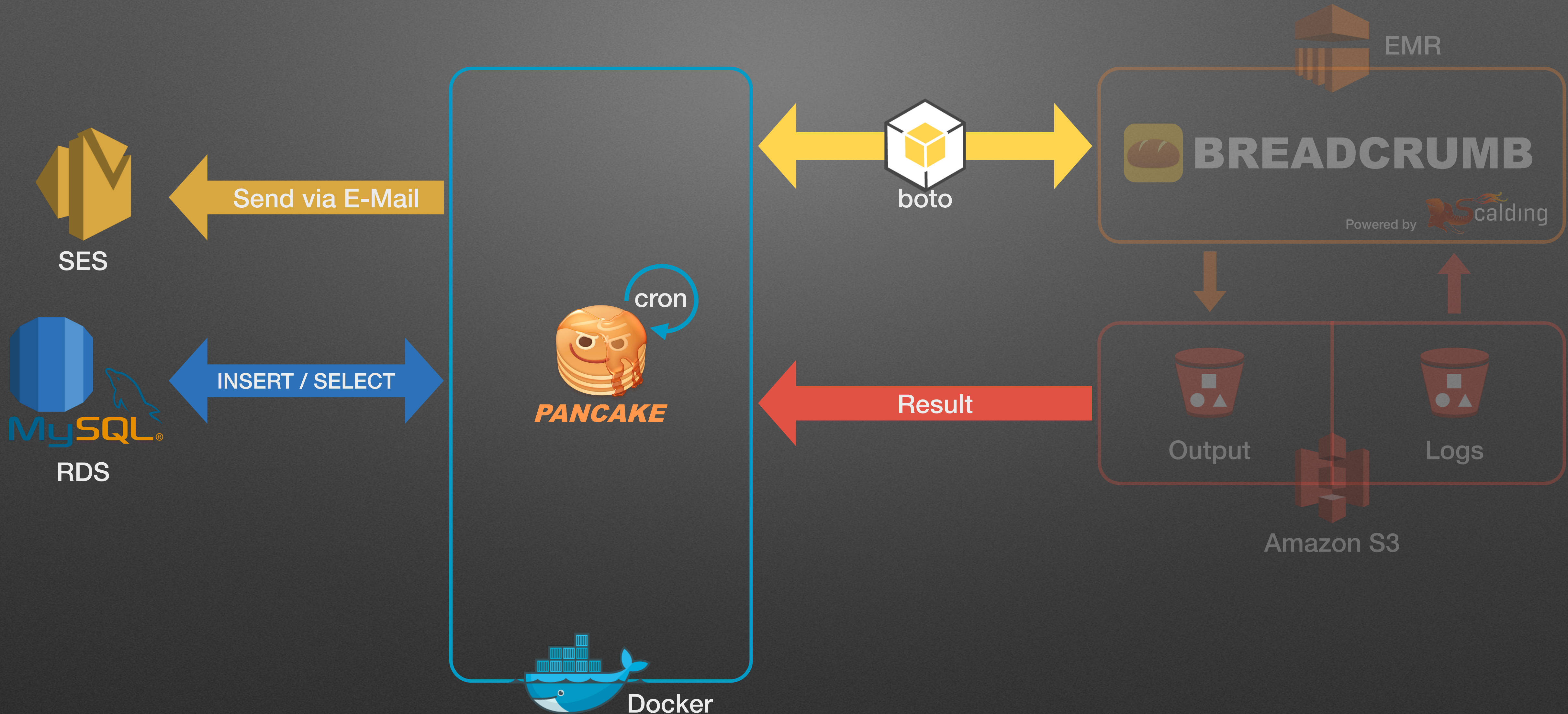
정기적인 통계가 필요하다!

2. 매일매일 데이터 받아보기

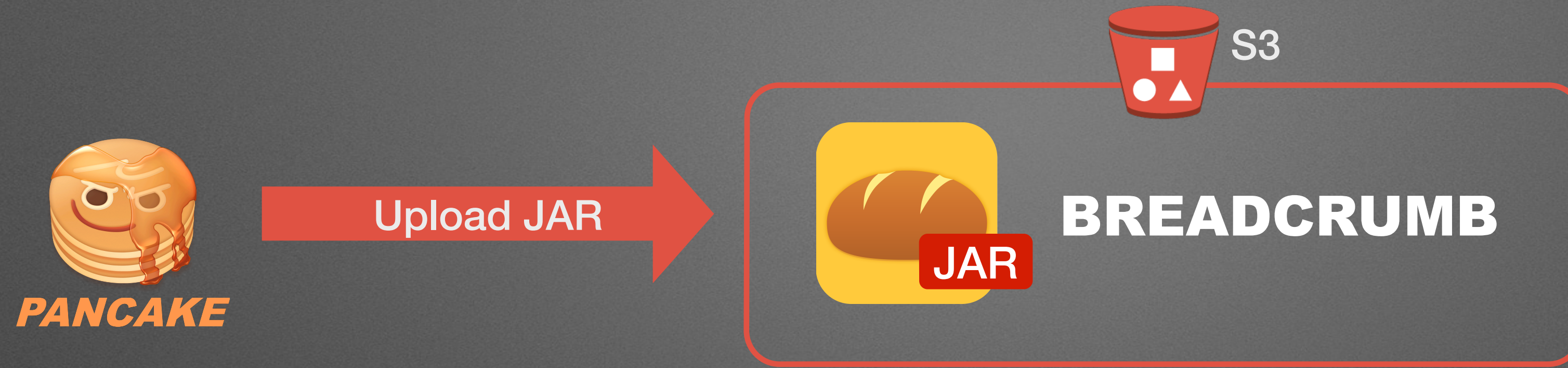
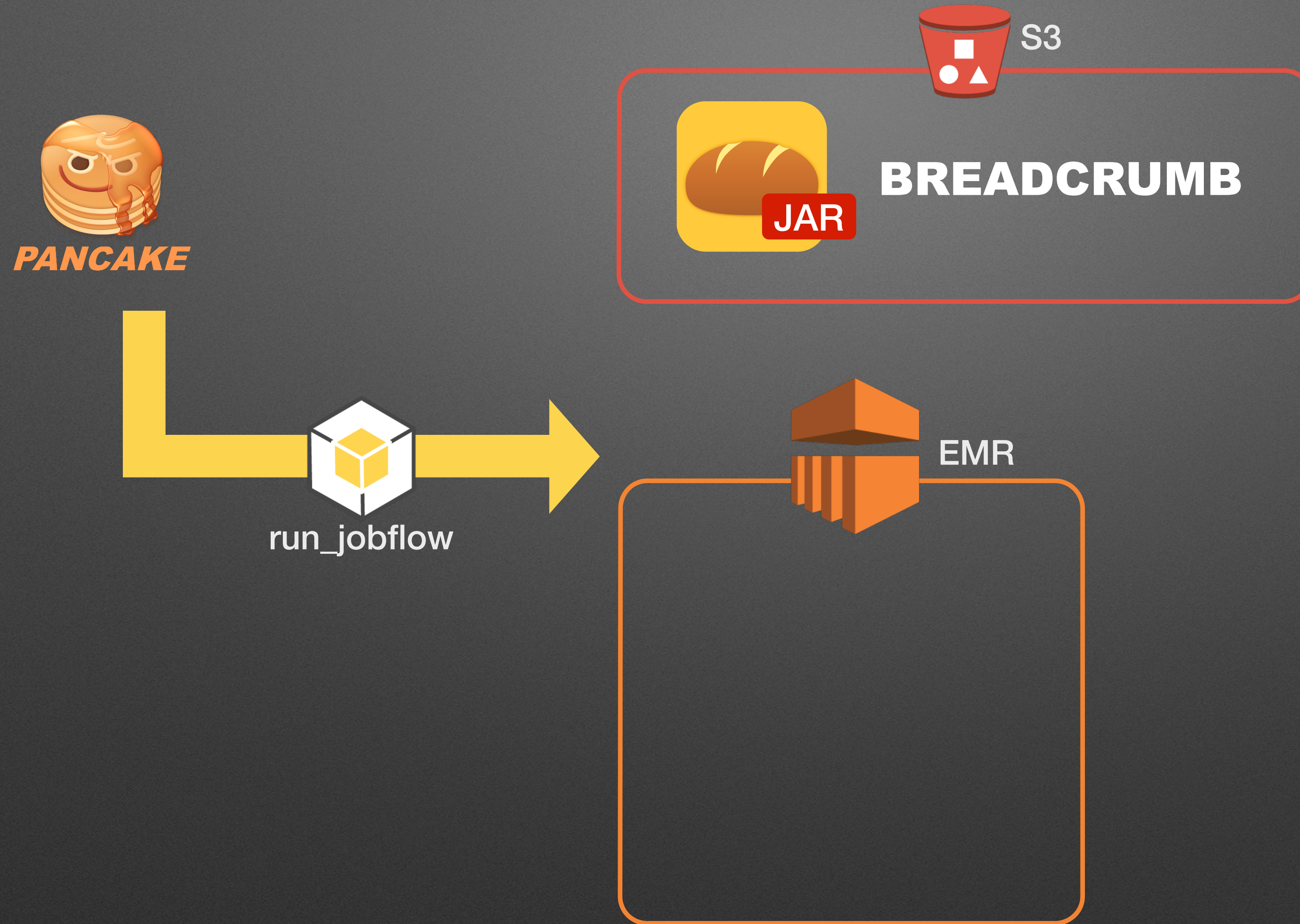
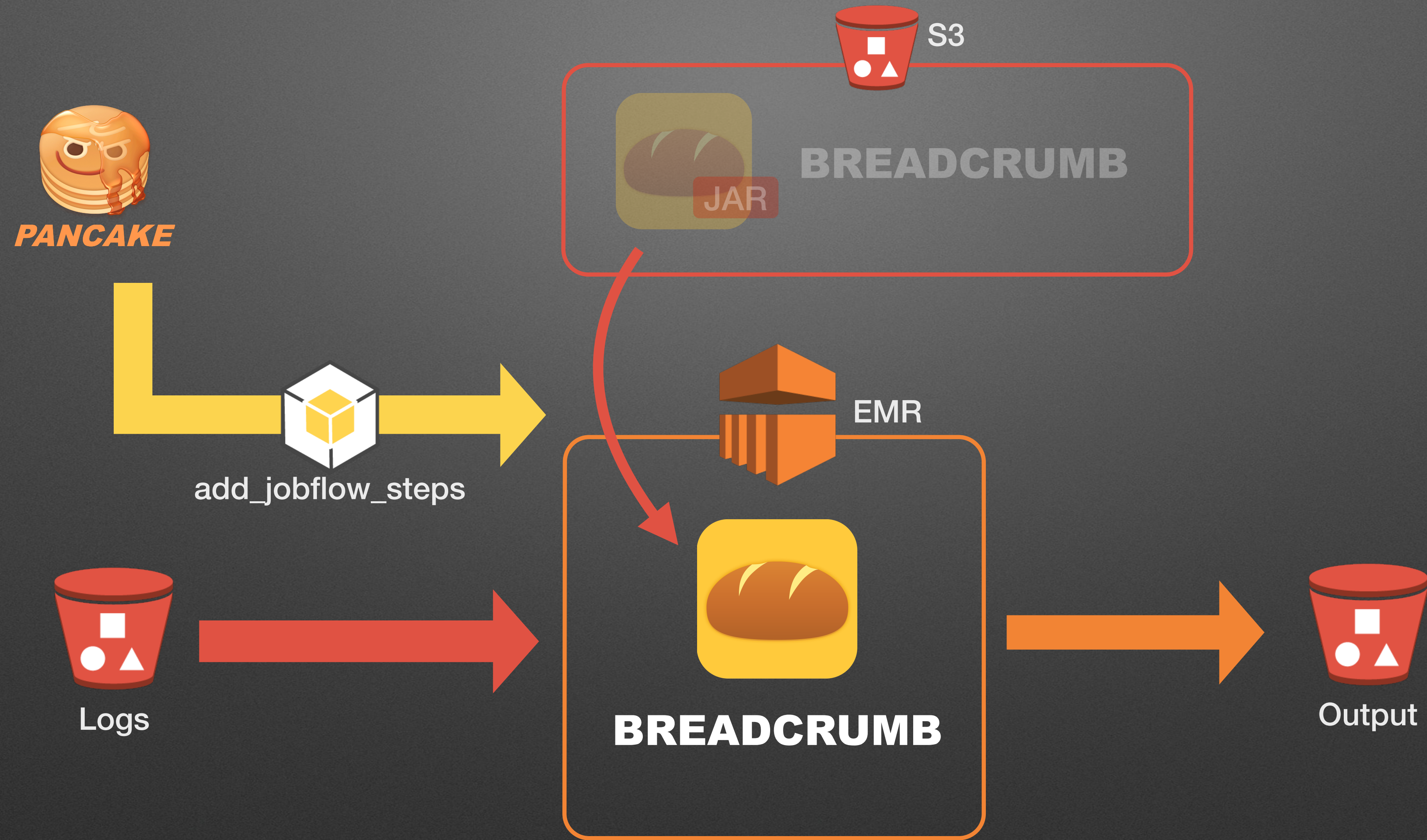
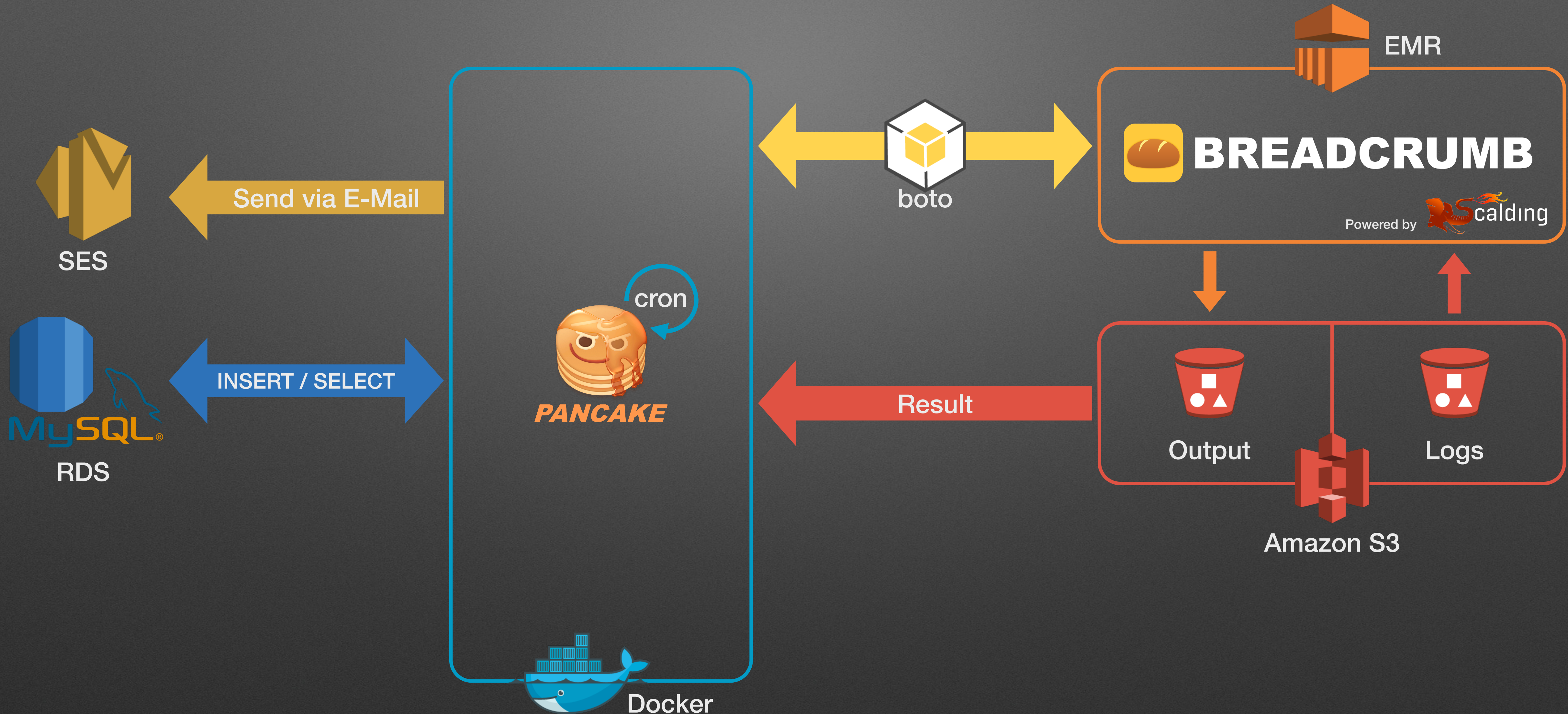
2. 매일매일 데이터 받아보기 - 목표

매일 아침 게임 통계를 받아보고 싶어요!

- □ 지표들의 변화 추이를 보고 싶어요!
- □ 쿠키별 보유량이 알고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

2-1. 매일 아침 게임 통계를 받아보고 싶어요!

날짜별로 정렬

Pancake의 요정 그저께
공유@ 2015-04-18 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 17.
공유@ 2015-04-17 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 16.
공유@ 2015-04-16 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 16.
공유@ 2015-04-16 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 15.
공유@ 2015-04-15 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 14.
공유@ 2015-04-14 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 13.
공유@ 2015-04-13 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 13.
공유@ 2015-04-13 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 12.
공유@ 2015-04-12 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 11.
공유@ 2015-04-11 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 10.
공유@ 2015-04-10 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정 2015. 4. 10.
공유@ 2015-04-10 지표 - Pancake
본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다. THIS MAIL CONTAINS IMPORTANT CLASSIF...

Pancake의 요정
받는 사람:
공유@ 2015-04-18 지표 - Pancake

본 메일은 중요한 사내 기밀을 담고 있습니다. 본 메일의 일부 혹은 전체를 외부에 제공하는 것은 엄격히 금지되어 있습니다.
THIS MAIL CONTAINS IMPORTANT CLASSIFIED INFORMATION. IT IS STRONGLY PROHIBITED TO PROVIDE PART OR ALL OF THIS MAIL.

Daily Active User : 2015-04-17

Daily New User : 2015-04-17

Daily Game Play User : 2015-04-17

Episode Daily Active User : 2015-04-17
Episode1 :
Episode2 :
Episode3 :
Episode4 :
Episode501 :

Daily Revenue: 2015-04-17

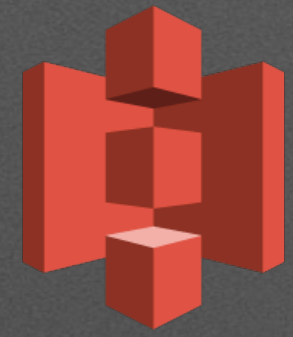
	$ USD	₩ KRW	전일 대비
Apple			%
Google			%
Total			

1 USD = 1085.32334 KRW (2015-04-18 09:00:49 기준)

Paid User

구매유저 수	구매유저 비율	전일대비

Gem Usages: 크리스탈 사용분포

순위	타입	사용량	비율(%)
1			%
2			%
3			%
4			%
5			%
6			%
7			%
8			%
9			%
10			%
11			%
12			%
13			%

Amazon S3

- 로그를 담아두기 위한 최고의 저장소
- 신뢰할 수 있는 스토리지
- Offset-Range를 지원하여 로그에서 필요한 부분만 가져오는 것도 가능

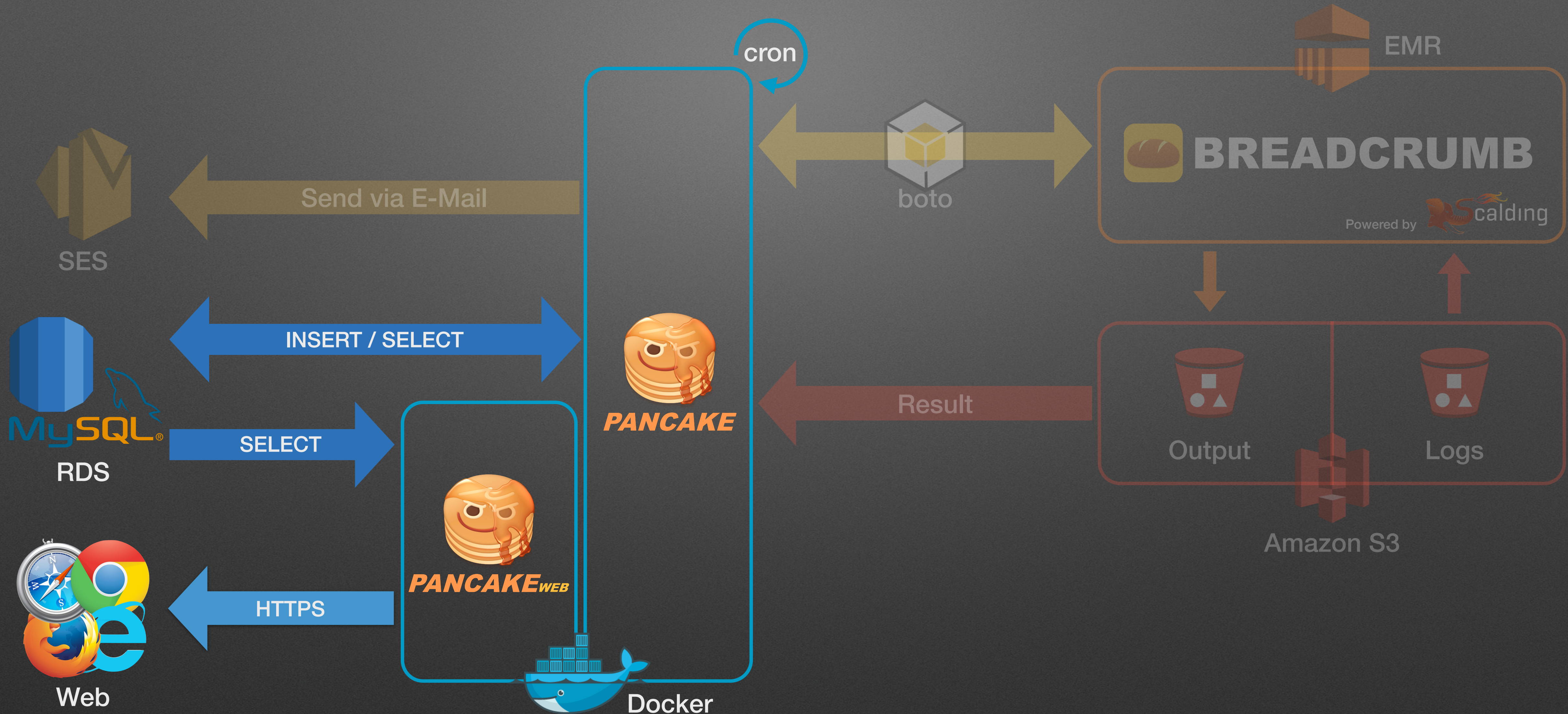
EMR

- 하루 한번 한시간정도 필요한 클러스터를 운영하기엔 최고
- Spot Instance로 사용하면 비용도 굉장히 저렴하다!
- S3에 Direct Access 가능 HDFS에 로그를 업로드할 필요가 없다

2. 매일매일 데이터 받아보기 - 목표

✓ 매일 아침 게임 통계를 받아보고 싶어요!

☐ 쿠키별 보유량이 알고 싶어요!

2-2. 지표들의 변화 추이를 보고싶어요!

2-2. 지표들의 변화 추이를 보고싶어요!

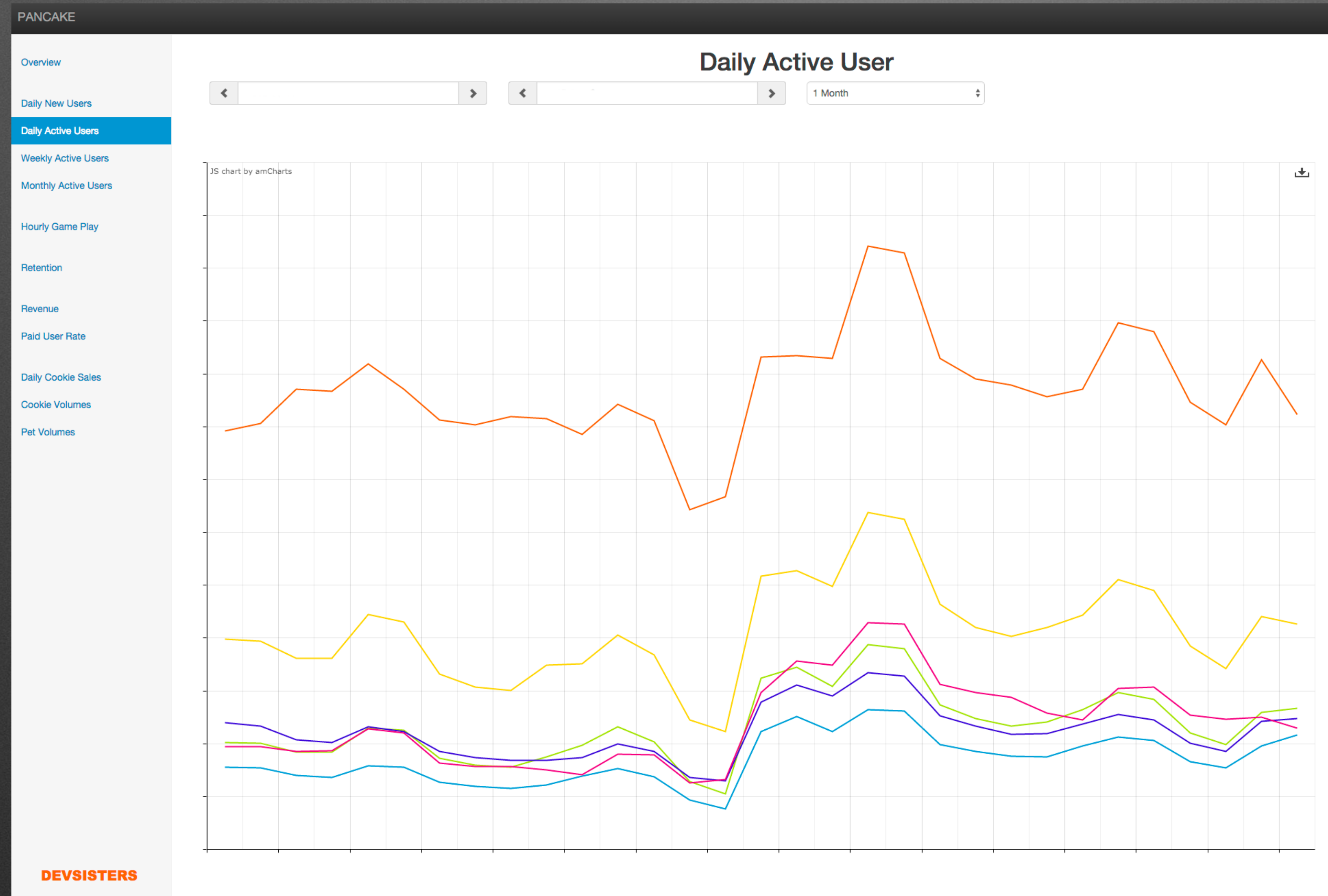
PANCAKE

- Overview
- Daily New Users
- Daily Active Users
- Weekly Active Users
- Monthly Active Users
- Hourly Game Play
- Retention
- Revenue
- Paid User Rate
- Daily Cookie Sales
- Cookie Volumes
- Pet Volumes

DEVSISTERS

Daily Active User

1 Month

JS chart by amCharts

2. 매일매일 데이터 받아보기 - 목표

✓ 매일 아침 게임 통계를 받아보고 싶어요!

✓ 지표들의 변화 추이를 보고 싶어요!

2-3. 유저의 쿠키 보유량이 알고 싶어요!

쿠키 보유량 = 쿠키의 총 판매량

2-3. 유저의 쿠키 보유량이 알고 싶어요!

쿠키 보유량 = 쿠키의 총 판매량

출시 이후 모든 판매로그를 분석..?

2-3. 유저의 쿠키 보유량이 알고 싶어요!

쿠키 보유량 = 쿠키의 총 판매량

출시 이후 모든 판매로그를 분석..?

2-3. 유저의 쿠키 보유량이 알고 싶어요!

유저데이터를 담고 있는

View를 활용하자!

2-3. 유저의 쿠키 보유량이 알고 싶어요!

Documents로부터 MapReduce를 통해 원하는 형태의 데이터로 변환하는 기능
Mongo의 Aggregate, MapReduce와 비슷

2-3. 유저의 쿠키 보유량이 알고 싶어요!

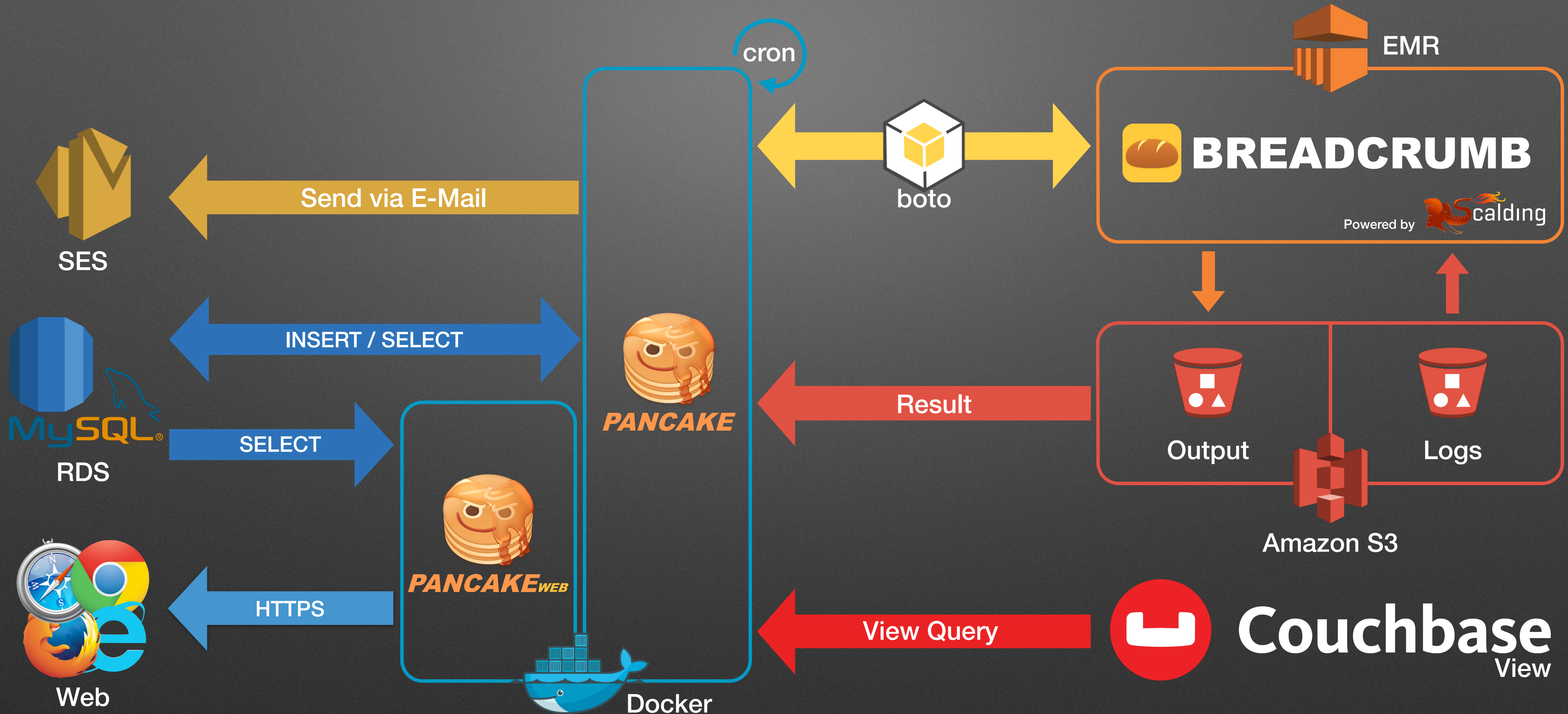
Documents로부터 MapReduce를 통해 원하는 형태의 데이터로 변환하는 기능
Mongo의 Aggregate, MapReduce와 비슷

유저 정보에 들어있는 쿠키 정보를 쿠키별 카운트로 변환!

[용감한 쿠키, 좀비맛 쿠키, 천사맛 쿠키]
[용감한 쿠키, 천사맛 쿠키, 악마맛 쿠키]
[용감한 쿠키, 공주맛 쿠키, 악마맛 쿠키]

MapReduce →

용감한 쿠키 : 3
천사맛 쿠키 : 2
악마맛 쿠키 : 2
공주맛 쿠키 : 1

2. 매일매일 데이터 받아보기 - 완성!

2. 매일매일 데이터 받아보기 - 완성!

✓ 매일 아침 전날의 게임 통계를 받아보고 싶어요!

✓ 지표들의 변화 추이를 보고싶어요!

✓ 쿠키별 보유량이 알고 싶어요!

3. 성과 - 로그분석시스템 Breadcrumb

- ✓ 분석에 필요한 로그만을 가져오는 분석 시스템
- ✓ Scalding 기반으로 쉽고 빠른 개발
- ✓ 로컬에서 샘플링된 유저집합으로 개발한 이후 동일하게 전체 데이터에 적용

- ✓ 현재까지는 당일의 데이터는 조회가 불가

  → 당일 로그의 분석도 가능한 시스템 개발 완료

3. 성과 - Daily Report Pancake

- ✓ 하루에 한번 자동화된 로그분석과 보고서
- ✓ 일회성 분석으로 작성한 코드도 곧바로 정기분석에 추가
- ✓ 여러 데이터 소스로부터 다양한 분석
- ✓ 게임 성과 분석 뿐 아니라 어뷰징 분석도 정기적으로 수행

- ✓ Breadcrumb의 실시간 로그분석에 대응한 버전을 준비 중

우리는 뛰어난 당신을 원하고 있습니다

http://www.devsisters.com/jobs/

서버개발팀
정창제
changje@devsisters.com